Acknowledgments

The authors thank Dr. I. R. Lehman for the gift of endonuclease I and Dr. J. G. Moffatt for the gift of several nucleosides and nucleotides.

References

Atkinson, M. R., Deutscher, M. P., Kornberg, A., Russell, A. F., and Moffatt, J. G. (1969), *Biochemistry 8,* 4897.

Byars, N., and Kidson, C. (1970), *Nature (London) 226,* 648.

Doering, A. M., Jansen, M., and Cohen, S. S. (1966), *J. Bacteriol. 92,* 565.

Haskell, E. H., and Davern, C. I. (1969), *Proc. Natl. Acad. Sci. U.S.A. 64,* 1065.

Josse, J., Kaiser, A. D., and Kornberg, A. (1961), *J. Biol. Chem. 236,* 864.

Khorana, H. G., and Vizsolyi, J. P. (1961), *J. Am. Chem. Soc. 83,* 675.

Kidson, C. (1966), *J. Mol. Biol. 71,* 1.

Kidson, C., Chen, P., Edwards, A., and Mansbridge, J. N. (1975), in The Eukaryote Chromosome, Peacock, J., and Brock, R., Ed. (in press).

Lehman, I. R. (1960), *J. Biol. Chem. 235,* 1479.

Lehman, I. R., Bessman, M. J., Simms, E. S., and Kornberg, A. (1958), *J. Biol. Chem. 233,* 163.

Pfitzner, K. E., and Moffatt, J. G. (1964), *J. Org. Chem. 29,* 1508.

Sugino, A., Hirose, S., and Okazaki, R. (1972), *Proc. Natl. Acad. Sci. U.S.A. 69,* 1863.

A Comparison of the Binding to Polynucleotides of Complementary and Noncomplementary Oligonucleotides†

James B. Lewis,‡ Lawrence F. Brass,§ and Paul Doty*

ABSTRACT: The binding of oligonucleotides to synthetic polynucleotides has been studied as a control for investigations of the binding of oligonucleotides to natural RNA molecules. Only combinations that involved A-U, G-C, and G-G pairs were found to be significantly stable under the experimental conditions used here. The stability of the oligomer–polymer pairing increased with the length of the region paired and with its G + C content. Further, some different sequence isomers of the same G + C content exhibited quite different binding constants. This variability is consistent with certain sequence differences in the double-strand stacking interactions stabilizing the oligomer–polymer association. Oligomer binding was also shown to depend upon the identity of the polymer residues neighboring the binding site, indicating the effect upon oligomer binding of small changes in the single-strand conformation of the binding site. These observations validate the criteria that allow one to decide if an observed association constant of an oligomer to an RNA molecule reflects a complete complementarity between the two or not. This improves the basis for using oligonucleotide binding constants to RNA of known sequence to map secondary structure.

Recent reports (Uhlenbeck et al., 1970; Lewis and Doty, 1970; Hogenauer, 1970; Danchin and Grunberg-Manago, 1970; Uhlenbeck, 1972; Schimmel et al., 1972) have used the binding of short oligonucleotides to RNA to investigate the secondary structure of small RNAs of known sequence. The supposition is that only those parts of the RNA sequence that are single stranded and available for base pairing will bind complementary oligonucleotides. Thus the sequence of the regions of the RNA chain that are single-stranded and available can be deduced from the sequences of the oligonucleotides that are found to bind to the RNA by assuming that the binding site for the oligonucleotide on the RNA molecule is the Watson–Crick antisequence of the oligomer. For this assumption to be valid requires first that only Watson–Crick specific interactions contribute significantly to the negative free energy of binding. Second, it is necessary that the binding constant for any part of an oligonucleotide sequence be sufficiently less than the binding constant for the entire sequence that the single-stranded region of the RNA can be identified as the complement of the entire oligonucleotide.

To determine if these two conditions are met, we have studied the binding of oligonucleotides to synthetic polynucleotides. The frequency with which various site sequences occur can be calculated on the basis of the random incorporation of the nucleotides into the polymer (Grunberg-Manago, 1963). The effect of secondary structure masking some of these sites, as is expected to occur with natural RNAs, should be eliminated by using only those bases in a given polymer that will not pair to each other. It will be seen, however, that in a few cases secondary structure still persists, and this structure must be considered in interpreting experiments with these few polymers. Each of the effects mentioned above can be studied by systematically varying the sequence of the oligomers used and the composition of the polymer.

† From the Department of Biochemistry and Molecular Biology, Harvard University, Cambridge, Massachusetts 02138. *Received January 6, 1975.* This work was supported by National Science Foundation Grant GB-27443 and by a National Science Foundation Fellowship to J.B.L.

‡ Present address: Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724.

§ Present address: Department of Biochemistry Case Western Reserve, Cleveland, Ohio 44118.

Materials and Methods

Synthesis of Polynucleotides. Homopolynucleotides used were purchased from Miles Chemical Company or from Calbiochem Company. Polynucleotides containing two different nucleotide residues were synthesized as described by Uhlenbeck (1969) using primer independent polynucleotide phosphorylase generously donated by Drs. O. Uhlenbeck and F. Martin. It was expected that the base composition of the polymers would be similar to the ratio of nucleoside diphosphates used in the preparation. The expectation was confirmed by determining the base composition by the method of Comb and Katz (1963). Concurrently, the hypochromicity, and thus the average extinction coefficient per residue of the polymer, was determined by measuring the absorbance of the polymer before and after alkaline hydrolysis.

Synthesis of Oligonucleotides. Tritium-labeled oligonucleotides were prepared as previously described (Uhlenbeck et al., 1970). A more detailed discussion of the synthesis of tritium-labeled tetramers is available (Lewis, 1971).

Measurement of Binding of Oligomer to Polymer. The binding of oligonucleotides to polynucleotides was measured by equilibrium dialysis as previously described (Uhlenbeck et al., 1970), with one modification. Unlike 5S RNA and tRNA, synthetic polynucleotides often aggregate in the presence of magnesium (Sander and Ts'o, 1971). This aggregation is so great that the binding of oligonucleotides to polynucleotides in the presence of 10 m*M* magnesium is almost independent of polynucleotide concentration. Thus the equilibration conditions used here are: 1 *M* NaCl, 5 m*M* EDTA, and 10 m*M* cacodylate (pH 7.2) at 0°. Poly(C) and poly(U) were sometimes used at concentrations up to 3 mg/ml, but the other polynucleotides were used at concentrations of less than 1.5 mg/ml. That there is little polynucleotide aggregation under these conditions was confirmed by finding that the ratio of bound to free oligomer is strictly proportional to the polynucleotide concentration within these concentration ranges.

Calculation of the Binding Constant (K). If K is the association constant of an oligomer to a single complementary site on the polynucleotide, R is the ratio of oligomer radioactivity (cpm) in the chamber containing the polynucleotide to that in the chamber without, and [S] the concentration of complementary sites, then the relationship $R = 1 + K[S]$ will hold as long as the site concentration greatly exceeds the oligomer concentration. The oligomer concentration in these experiments ranged from 0.02 to 1.4 μM while the site concentrations ranged from 15 μM to 10 m*M*.

The site concentration [S] can be related to the absorbance in the chamber containing the polynucleotide by the relation $[S] = (A_{260}/E_{res})f$, where E_{res} is the average molar extinction coefficient per residue and f is the fraction of residues in a given sequence. Since all of the polynucleotides studied contain no more than two different bases incorporated into a random sequence, the fraction (f) of residues in a given sequence is $f = (a)^m(1 - a)^{n-m}$ where a is the fraction of residues in the polymer of type A, and m is the number of times nucleotide residue A occurs in the site sequence n residues long. Thus

$$K = (R - 1)E_{res}/fA_{260}$$

In these experiments values of $(R - 1) < 0.02$ are experimentally indistinguishable from the case of no binding. Except for the weak dimer binding of K's reported here are all based on $(R - 1)$ values much greater than 0.02. Multiple determinations were performed for one-fourth of the approximately 116 binding constants reported here: the average range of values was approximately 14% of the mean value.

Table I: Effect of Interoligomer Interaction upon the Binding of Oligomers.

Complex	Polymer Residue Concn (m*M*)	Oligomer Concn (μM)	f	Site Concn/ Oligomer Concn	K (M^{-1})
Poly(U) + AAAA	7.83	0.026	1.000	306,000	3,000
	3.31	0.029	1.000	116,000	7,690
	1.06	0.037	1.000	28,700	16,800
	0.69	0.042	1.000	16,200	20,400
Poly(C) + GG	10.00	0.5	1.000	20,000	95
	9.82	0.7	1.000	14,000	100
	4.36	1.2	1.000	3,600	107
	1.73	1.4	1.000	1,200	104

Results

The Specificity of Oligomer–Polymer Binding. To determine the fidelity with which an oligomer will pair with its Watson–Crick partner, the binding of homooligomers of the four bases A, C, G, and U to the various homopolynucleotides was studied. Since only one type of base pair is involved in each experiment, the stabilities of each of the ten possible pairs between these four bases can be compared. Unlike other studies of the complexes formed between homopolymers (Michelson et al., 1967) or between oligomers and polymers (Uhlenbeck, 1969; Michelson and Monny, 1967), which involve stoichiometric quantities of the components, these experiments were done with a large excess (10^2–10^4-fold in nucleotide concentration) of polymer. This avoidance of concentrations anywhere near stoichiometric was necessary since in that range the binding of oligomer to polymer is greatly stabilized by interoligomer stacking with the result that the association of oligomers to polymer becomes highly cooperative and the simple site binding of oligomers, which we wish to study, is not displayed.

Binding under such cooperative conditions will be much stronger than the association of an oligomer to a natural RNA, where adjacent sites are not likely to exist. The experiments reported here use a large excess of binding sites to oligomer so that the entropy gained by binding to the numerous nonadjacent sites would overcome the stabilization gained by interoligomer stacking. In this way the K obtained should correspond to the binding of an oligomer to a natural RNA.

This approach was largely successful as attested by the apparent value of K being independent of the ratio of polymer to oligomer. The only failure occurred with poly(U) + A_4. The data shown in Table I demonstrate that the apparent value of K continues to fall as the polymer concentration is increased with no convergence in evidence within the accessible concentration range. By contrast the apparent value of K for poly(C) plus G_2 remained constant when the polymer concentration was varied: this was the usual result.

A larger collection of data is presented in Table II showing the range of association constants for homopolymer–homooligomer interaction. Data for two copolymers are included as well. It was necessary to substitute poly($U_{1.2}$,G) since poly(G) was not soluble in 1 *M* NaCl. The oligomers

Table II: Relative Stabilities of Different Types of Base Pairs.

Oligomer	Polymer	Residue Concn of Polymer (mM)	f	R	K (M^{-1})
AAAA	Poly(A)	4.98	1.000	1.02	4
	Poly(C)	10.36	1.000	1.00	<2
CCCC	Poly(A)	5.42	1.000	1.00	4
	Poly(C)	10.52	1.000	1.00	<2
	Poly(U)	8.41	1.000	0.99	<2
	Poly($U_{1,2}$,G)	5.64	0.037[a]	7.04	28,900
GG	Poly(A)	5.12	1.000	1.06	12
	Poly(U)	7.61	1.000	1.02	3
	Poly($U_{1,2}$,G)	6.39	0.189[b]	4.29	2720
		5.25	0.189	3.42	2440
	Poly(C)	9.82	1.000	2.00	100
UUUUU	Poly(C)	10.18	1.000	0.98	<2
	Poly(U)	7.74	1.000	1.01	<2
	Poly($U_{1,6}$,C)	6.86	0.092[c]	1.04	60
	Poly($U_{1,2}$,G)	5.01	0.016[d]	1.00	<250
	Poly(A)	5.17	1.000	2.83	350
UUUU	Poly(A)	5.48	1.000	1.27	49

[a] Expected frequency of the sequence . . .GGGG. . . . [b] Expected frequency of the sequence . . .GG. . . . [c] Expected frequency of the sequence . . .UUUUU. . . . [d] Expected frequency of the sequence . . .GGGGG. . . . The very low frequency of GGGGG in this polymer makes the measurement of K for this interaction very imprecise.

Table III: The Binding of GG to CC Sequences in Different Polymers.

Polymer	Residue Concn of Polymer (mM)	f	R	K (M^{-1})
Poly(C)				102[a]
Poly($C_{6,1}$,A)	9.36	0.741	3.77	400
	9.64	0.741	4.13	438
	4.18	0.741	2.36	440
Poly($C_{1,9}$,A)	6.58	0.429	3.41	854
	5.58	0.429	3.82	1177
Poly($A_{1,2}$,C)	3.42	0.206	2.68	2385
	2.68	0.206	2.23	2225
Poly($A_{1,5}$,C)	3.71	0.159	2.88	3190
	3.97	0.159	2.35	2145
Poly($A_{2,4}$,C)	4.23	0.086	2.66	4570
	4.83	0.086	2.90	4570
Poly($A_{4,6}$,C)	9.97	0.033	2.86	5770
	10.43	0.033	2.97	5720
	5.06	0.033	2.09	6040
Poly($C_{2,6}$,U)	8.74	0.514	2.10	242
	8.38	0.514	2.05	244
Poly($U_{1,6}$,C)	6.32	0.146	1.71	770
	6.65	0.146	1.65	670

[a] Average of four determinations in Table I.

Table IV: Effect of Oligomer Composition upon the Differential Binding of the Oligomer to Polymers of Different Compositions.

Oligomer	K with Poly($C_{6,1}$,A)	K with Poly($A_{4,6}$,C)	K with Poly($A_{4,6}$,C)/ K with Poly($C_{6,1}$,A)
GG	426[a]	5,840[a]	14
GGU	11,700	82,000	6.7
	15,900	104,000	
GUG	5,190	37,900	7.5
	5,390	41,400	
GUU	652	1,200	1.8
GUUU	6,960	9,080	1.3

[a] Average from data in Table III.

varied in length from two to five residues to compensate for the very different stabilities of the Watson-Crick complexes formed by oligomers of different composition. In general, the oligomer length was chosen to make the complexes with the Watson-Crick complementary polymer strong enough so that the absence of interaction of that oligomer with other polymers would be meaningful. Thus a pentamer (U_5) was required to study binding in the case of U, while a dimer (G_2) was sufficiently long for comparable studies of G. From an examination of the results it is clear that only A·U, G·C, and G·G interactions are important.

However, it is not possible to make a direct comparison of the strength of G·C and G·C interactions from the data in Table II because the apparent K for GG binding to poly($U_{1,2}$,G) is strongly influenced by the considerable secondary structure of poly($U_{1,2}$,G), and by the presumed ability of GG to form large aggregates with the high concentrations of G residues in the polymer. The G·C interaction is about 1000 times as large as the A·U interaction, which in turn is much larger than any other interaction except perhaps G·U. No binding of U_5 to poly($U_{1,2}$,G) was observed, but because of the very small f for G_5 sites, we cannot exclude a binding of as much as 70% that of U_5 to poly(A).

Dependence of the Binding Constant (K) upon the Environment of the Site. To ascertain the dependence of the binding constant upon the sequence adjacent to the site, the binding of GG was measured to poly(C) and to eight copolymers containing A and C or U and C in varying proportions. By varying the compositions of the polymers, and thus the identity of the residues neighboring the binding site, the environment of the binding site can be changed. The variation in K for the binding of GG that occurs with changes in polymer composition is shown in Table III. It can be seen that all of the copolymers studied bind GG better than does poly(C), with the (A,C) copolymers providing better binding than the (U,C) copolymers. Neighboring A residues seem to provide a better environment for binding to the sequence CC than to neighboring U residues, presumably due to the fact that A residues stack much better than do U residues. Neighboring C residues provide a very poor environment for binding, even compared to U residues, which stack very little. This unexpected result would appear to indicate that C residues in poly(C) were stacked in a way unfavorable to oligomer binding.

If some aspect of poly(C) stacking diminishes association constants for G·C pairing the effect should be lessened by interrupting the sequences adjacent to the binding sites. To test this effect the binding constants for various (G,U) oligomers to two different copolymers of A and C were measured: the results are listed in Table IV. It can be seen that the effect of the composition of the polymer upon K decreases sharply as the number of C's in the antisequence decreases. As the amount of U in the oligomer is increased, the binding site for the oligomer increases in A content and becomes less typical of the average environment of a C-rich copolymer and more like that of an A-rich copolymer.

To be sure that changes in the amount of oligomer bound are due to changes in the environment of the site and not to changes in the number of binding sites, it is necessary that

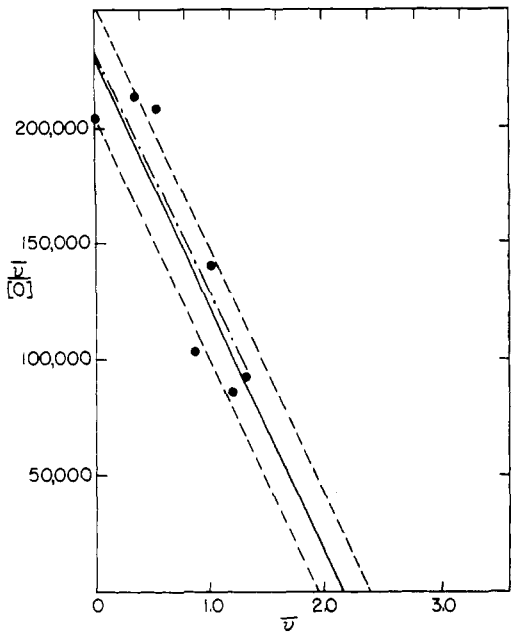

FIGURE 1: Scatchard plot of G-G-U binding to poly($A_{4.5}$,C). Least-squares straight line ± one standard deviation fitted to the equation $\bar{\nu}/[O] = K(n - \bar{\nu})$. [O] is the concentration of free oligonucleotide. Given the specific activity of the oligonucleotide, [O] is calculated from the radioactivity at equilibrium in the side of the dialysis chamber without polynucleotide. $\bar{\nu}$ is the total average number of sites occupied per polymer molecule. The difference in radioactivity between the two sides of the dialysis chamber gives the total number of sites on the polynucleotide occupied by oligonucleotide. $\bar{\nu}$ is obtained by dividing this result by the total concentration of polynucleotide. For the purpose of calculation, the total polynucleotide concentration is arbitrarily set at 0.01 times the residue concentration of polynucleotide. Thus $n = 100f$. The values of f and k calculated from the least-squares straight line are $f = 0.022 \pm 0.002$ and $K = 105{,}000$. To calculate the binding curve expected if one includes the possibility of G-G-U binding as the dimer G-G to C-C sequences: $\bar{\nu} = [n_1K_1[O]/(1 + K_1[O])] + [n_2K_2[O]/(1 + K_2[O])]$, where n_1 and K are equal to n and K determined above and n_2 and K_2 characterize the binding of G-G-U to C-C sequences that are part of C-C-C rather than A-C-C sequences. Thus $n_2 = 0.6$ ($f = 0.006$) and $K_2 \simeq 6000$, which is the K of G-G binding to poly($A_{4.5}$,C). However, since 82% of the C-C sites for G-G binding to poly($A_{4.5}$,C) are in A-C-C rather than C-C-C sequences, and since a long sequence of C residues is unfavorable for G-G binding, K_2 is probably much less than 6000. On the other hand, as is shown below, a nonbinding U residue on the 3′ end of an oligonucleotide can slightly increase the K characterizing the binding of the remainder of the sequence. This effect is expected to be less than the first, however. Thus, the calculated curve (— ··) probably represents the maximum deviation from linearity expected from the inclusion of the binding of G-G-U as the dimer G-G.

f, the fraction of nucleotides in the polymer that are located in complementary sites, be known accurately. Thus, f was determined directly in one case. This was done by studying the binding of GGU to poly($A_{4.6}$,C) over a range of concentrations (Tanford, 1961). GGU was used instead of GG because its higher K (see below) permitted the range of oligonucleotide concentrations used to be lower than had GG been used, minimizing the effect of oligo(G) aggregation at high concentrations.

The data are plotted in the Scatchard convention in Figure 1. The result shows a linear behavior corresponding to $f = 0.022 \pm 0.002$, slightly lower than the predicted result of 0.027. If the polymer were totally nonrandom, i.e., the C residues were all adjacent in the polymer, f would have been more than five times higher than predicted. Thus our assumption of random copolymerization seems justified.

In summary, it has been demonstrated that the association constant for the binding of an oligomer to a polymer depends not only on the sequence of the binding site, but also on the identity of the residues neighboring the site. Two components of this contribution have been identified. (a) If the neighboring residues stack poorly (i.e., are U's), the site binding will be lower than if the neighboring residues stack well. (b) Long sequences of C's are difficult to put into a double helical array.

Table V: Dependence of K upon Oligomer Length for Several Series of Homologous Oligomers.

Oligomer	Polymer	Residue Concn of Polymer (mM)	R	f	K (M^{-1})
AAA	Poly($U_{1,6}$,C)	2.65	3.48	0.236	3,850
AAAA	Poly($U_{1,6}$,C)	2.70	15.26	0.146	36,200
CC	Poly($U_{1,2}$,G)	4.94	1.04	0.189	43[a]
CCC	Poly($U_{1,2}$,G)	5.04	1.83	0.085	1,940[a]
CCCC	Poly($U_{1,2}$,G)	5.64	7.04	0.037	28,900[a]
UUU	Poly(A)	4.83	1.01	1.000	4
UUUU		5.48	1.27	1.000	49
UUUUU		5.17	2.83	1.000	353
GG	Poly(C)				102[b]
**GGG		4.51	3.40	1.000	533[c]
		4.66	5.73	1.000	1,020[c]
GG	Poly($U_{1,2}$,G)				2,580[a,d]
GGG	Poly($U_{1,2}$,G)	2.29	5.10	0.085	21,100[a]
	Poly($U_{1,2}$,G)	2.34	6.01	0.085	25,200[a]
GG	Poly($C_{6,1}$,A)				426[e]
GGG	Poly($C_{6,1}$,A)	3.86	5.89	0.638	1,990
	Poly($C_{6,1}$,A)	3.93	7.44	0.638	2,550
GG	Poly($A_{4,6}$,C)				5,840[e]
GGG	Poly($A_{4,6}$,C)	2.57	7.97	0.006	452,000
	Poly($A_{4,6}$,C)	2.64	8.90	0.006	500,000
GG	Poly($C_{2,6}$,U)				243[e]
GGG	Poly($C_{2,6}$,U)	3.66	6.58	0.369	4,880
	Poly($C_{2,6}$,U)	3.40	8.03	0.369	5,610
GG	Poly($U_{1,6}$,C)				720[e]
	Poly($U_{1,6}$,C)	2.88	6.05	0.056	31,400
	Poly($U_{1,6}$,C)	3.01	6.25	0.056	31,100

[a] K not corrected for the secondary structure of Poly($U_{1,2}$,G). [b] Average of four determinations in Table I. [c] GGG equilibrates very slowly due to the tendency of G containing oligomers to aggregate. Consequently all determinations of the association constant of GGG were done in two ways to check that equilibrium was actually obtained. In one method, the equilibrium dialysis experiment was set up in the usual way, with the oligonucleotide initially added to the chamber without polynucleotide. In the other, the oligonucleotide was added initially to the same chamber as the polynucleotide. The two methods would of course give the same result at equilibrium. In most cases equilibrium was finally reached, although in this particular case equilibration had not yet been completed by 17 days. Thus, the value of K is probably intermediate between these two values. [d] Average of two determination in Table II. [e] Average of determinations in Table III.

The Association Constant Increases Sharply with the Length of the Region Binding and Depends Less Markedly on Nonbinding Nucleotides in the Oligomer. To aid in distinguishing the pairing of the entire sequence of an oligomer to RNA from the pairing of only part of that sequence, the extent to which adding an additional residue to an oligomer affects K was determined for a number of cases. Naturally, if this added residue can also pair with the polymer, the value of K would be expected to increase as was demonstrated in studies with self-complementary oligomers (Martin and Uhlenbeck, 1971). Such effects are seen in our data on the binding of oligomers to polymers, which is assembled in Table V. The data cover both Watson-Crick pairs and G·G interactions. The increase in the value of K is

Table VI: Effect of a Nonbinding Base(s) upon K for GG Binding to Various Polymers.

Oligomer	Polymer	K (M^{-1})[a-c]
GG	Poly(C)	102
AGG	Poly(C)	76, 81, 87, 97 (85)
AAGG	Poly(C)	86
CGG	Poly(C)	84, 90, 104 (93)
GGA	Poly(C)	115
GGC	Poly(C)	140
GGU	Poly(C)	435 +
GGUU	Poly(C)	336 +
UGG	Poly(C)	36, 46, (41) −
GG	Poly($C_{6,1}$,A)	426
AGG	Poly($C_{6,1}$,A)	284 −
CGG	Poly($C_{6,1}$,A)	236 −
GGA	Poly($C_{6,1}$,A)	321
GGC	Poly($C_{6,1}$,A)	512
GG	Poly($A_{4,6}$,C)	5,840
AGG	Poly($A_{4,6}$,C)	4,660
CGG	Poly($A_{4,6}$,C)	2,780 −
GGA	Poly($A_{4,6}$,C)	5,520
GGC	Poly($A_{4,6}$,C)	7,290 +
AAGG	Poly($A_{4,6}$,C)	3,620 −
ACGG	Poly($A_{4,6}$,C)	8,060 +
AGGA	Poly($A_{4,6}$,C)	7,960 +
AGGC	Poly($A_{4,6}$,C)	7,680 +
CAGG	Poly($A_{4,6}$,C)	7,970 +
GGAA	Poly($A_{4,6}$,C)	6,010
GGCC	Poly($A_{4,6}$,C)	7,560 +
GG	Poly($C_{2,6}$,U)	243
CGG	Poly($C_{2,6}$,U)	248
GGC	Poly($C_{2,6}$,U)	397 +
GGU	Poly($C_{2,6}$,U)	980 ++
UGG	Poly($C_{2,6}$,U)	297
GG	Poly($U_{1,6}$,C)	720
CGG	Poly($U_{1,6}$,C)	622
GGC	Poly($U_{1,6}$,C)	876
GGU	Poly($U_{1,6}$,C)	1,440 +
UGG	Poly($U_{1,6}$,C)	1,060
GGCC	Poly($U_{1,6}$,C)	1,080
GGUU	Poly($U_{1,6}$,C)	2,370 +
UGGC	Poly($U_{1,6}$,C)	2,670 +
UGGU	Poly($U_{1,6}$,C)	2,110 +
UUGG	Poly($U_{1,6}$,C)	900

[a] All K's are calculated from "f" for binding to CC sequences in the polynucleotide. [b] The average of several determinations is given in parentheses. [c] Instances in which the additional nonbinding base decreases or increases K compared to K for GG binding to the polynucleotide are noted by − or +.

seen to vary from as little as a factor of 5 to as much as a factor of 100 in the various cases studied.

Having established that a strong chain-length dependence is operative for short oligomers, it is necessary to compare the effects of adding a base that can pair with the effect of adding a base that cannot. It is the difference between the effects on the value of K of these two cases that will permit discrimination between complete and incomplete base pairing between an oligomer and a potential site on an RNA molecule.

The effect upon the association of GG with various polymers of adding an additional nonbinding base(s) to the GG dimer is shown in Table VI. In 26 out of 41 cases, K is found to be essentially unaffected (less than ±35%), but the nonbinding base significantly lowers K in four cases and significantly raises it in 11 cases. The effect of the nonbinding base ranges from a factor of 2.5 decrease compared to the K of GG binding to the same polymer to a factor of four increase.

A careful examination of those cases in which K is significantly changed shows that these changes are almost always associated with a nonbinding pyrimidine. However, it is not true that the addition of a nonbinding pyrimidine always changes K. In five out of 13 cases, the addition of a single nonbinding pyrimidine had little effect on K.

One can also conclude that a nonbinding pyrimidine residue will generally raise the value of K if a second nonbinding pyrimidine is also present. Five out of the seven cases where two nonbinding pyrimidines were added showed large increases in K. Whether or not a nonbinding pyrimidine will change K in a particular instance seems to depend as much on the environment of the site as on the identity of the oligomer. For example, UGG binds to poly(C) much less well than does GG, but binds to poly(U,C) equally well or better than does GG. However, in those instances in which K is changed, the amount of change in K does seem to correlate with the identity of the pyrimidine. A nonbinding C never caused a change of more than 60%, while a nonbinding U occasionally caused as much as a fourfold change in K.

The effect of the nonbinding base also seems to correlate with the sequence of the nonbinding oligomer. All four cases in which K was lowered involved the addition of the nonbinding base to the 5′ end of the GG sequence. On the other hand, five of the six cases in which a single pyrimidine caused a substantial increase in K involved addition to the 3′ end of the dimer.

Having established the range of effects upon K of adding a *nonbinding* base to GG, we consider next the extent of increase in K to be expected from the addition of a *binding* base. The effect of adding a third G to the dimer GG was shown in Table V to be an increase in K of a factor of 100 for binding to poly($A_{4,6}$,C) and a factor of 40 for binding to poly($U_{1,6}$,C). We now examine the other trimers containing the sequence GG that are complementary to these two polymers. First, it must be noted, however, that, unlike the case of nonbinding trimers (i.e., GG plus a nonbinding base), calculating K for the binding trimers is not straightforward, because the polymer not only contains sites to which the entire trimer can bind, but also sites to which only the GG portion of the trimer can bind.

To minimize this ambiguity, only the binding of oligomers to polymers low in C is reported in Table VII. Therefore, the number of sites to which a trimer can only bind as GG will be small compared to those to which it can bind as a trimer. Thus we justify calculating K in Table VII by the naive approach, ignoring the binding of only the GG portion of the trimer sequence.

In all four cases, the trimer binds much better than does GG. Except for UGG, the increase in K compared to the binding of GG is much greater than the largest increase in K observed upon the addition of a nonbinding base to GG. To a polymer to which GGU can bind as a trimer, it binds

Table VII: Effect of Adding a Binding Base upon the Binding of GG to Various Sequences.

Oligomer	Polymer	f	K (M^{-1})
GG	Poly($A_{4,6}$,C)	0.033	5,840
GGU	Poly($A_{4,6}$,C)	0.027	90,000 ± 14,000[a]
UGG	Poly($A_{4,6}$,C)	0.027	15,100 ± 2,500[a]
GG	Poly($U_{1,6}$,C)	0.146	720
AGG	Poly($U_{1,6}$,C)	0.090	17,500 ± 1,100[a]
GGA	Poly($U_{1,6}$,C)	0.090	7,700

[a] Average ± deviation of two determinations.

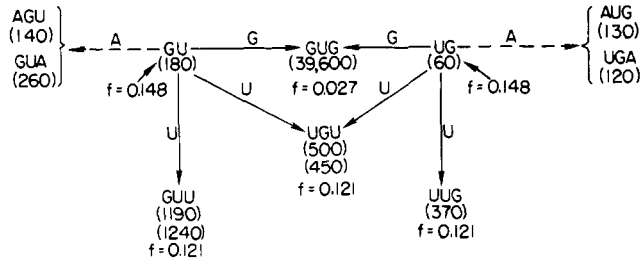

FIGURE 2: K for the binding to poly($A_{4.5}$,C) of trinucleotides containing the sequences G-U and U-G. The binding constant K for each oligonucleotide is given in parentheses. Solid arrows represent the addition of a binding base to the dinucleotide sequence; dashed arrows, the addition of a nonbinding base.

16 times better than does GG to the same polymer, while to a polymer to which it can only bind as a dimer, GGU binds only two to four times better than does GG. Thus GGU binds four to eight times better when it is binding as a trimer than when it is binding only the GG portion of its sequence. Thus, the large increase in K signifies that the oligomer is binding to an ACC rather than only to a CC sequence.

Since AGG and GGA do not bind significantly better than does GG when they are binding only part of their length, the difference between binding as a trimer and binding as a dimer is a factor of 20, even larger than with GGU. The binding of UGG is more ambiguous. When UGG binds as a trimer, it binds only 2.6 times better than does GG. Depending upon the environment of the binding site, it may bind as a dimer much less well than does GG, or as much as 50% better than GG. Consequently, the difference between binding as trimer and binding as dimer is relatively small so that it may be difficult to identify the sequence to which UGG is binding.

Adding a Binding Base to Dimers Other Than GG Also Causes a Much Larger Increase in K Than Adding a Nonbinding Base. To demonstrate that the addition of a nonbinding base to a dimer always causes a small change in K while the addition of a binding base causes a large increase, we examined several specific cases. The dimers GU and UG are expected to bind to poly(A,C), although very weakly because of their short length and intermediate G + C content. The K values of trimers formed by adding a binding base (G or U) to these trimers are displayed in Figure 2 along with those of the dimers and of the trimers formed by adding a nonbinding base (A). The oligomers GGU and UGG are not included because they have already been discussed in terms of the dimer GG. GGU and UGG clearly bind much better than do GU and UG, respectively, but the relevant comparison is with the binding of GG since GG binds much better than do GU and UG. Because of the low K values of the dimers and nonbinding trimers, the estimates of these K's are subject to an error in the range of ±30 due to the error of ±0.02 in determining R. It can be seen that the K's of the binding trimers vary over a wide range due to variations in G + C content, and to a lesser degree, variations in sequence within the same G + C content. However, the important observation is that for either dimer, the *smallest* increase in K due to the addition of a binding base is three times the *largest* increase due to the addition of a nonbinding base.

Adding a Binding Base to a Trimer Generally Causes a Much Larger Increase in K Than Does Adding a Nonbinding Base. We now consider the K value of selected tetramers formed by adding a binding or a nonbinding base to a binding trimer. The examples involve the binding of the trimers GUU, UUG, UGG, and GGU to poly(A,C), and these results are listed in Table VIII. Before the K values of the nonbinding and binding tetramers can be compared, it is necessary to indicate the way in which these K values are calculated.

Table VIII: Binding of Selected Tetramers to Poly(A,C)—82% A.

Oligomer	$(R-1)E/A_{260}$	f	K (M^{-1})
GUU	144; 150	0.121	1,190; 1,240
GUUC	160	0.121	1,320
GUUU	902	0.099	9,080
UUU	Too weak to detect		
GUUG	3,850	0.022	176,000
UUG	45	0.121	370
AGUU	134	0.121	1,100
UGUU	478	0.099	4,810
UGU	55; 61	0.121	450; 500
GGUU	4,900	0.022	225,000
GGU	2,040	0.027	76,400
UUG	45	0.121	370
UUGC	140	0.121	1,160
UUGG	2,440	0.022	112,000
UGG	344; 475	0.027	12,700; 17,600
UUUG	297	0.099	2,990
GUUG	Listed above		
UGU	55; 61	0.121	450; 500
AUGU	56	0.121	460
UGUU	Listed above		
UGG	344; 475	0.027	12,700; 17,600
UGGC	1,580	0.027	59,100
UGGU	4,570	0.022	209,000
GGU	2,040	0.027	76,400
CUGG	359	0.027	13,400
UUGG	Listed above		
GGU	2,040	0.027	76,400
AGGU	120; 130	0.027	4,450; 4,850
UGGU, GGUU	Listed above		
GUG	1,010; 1,100	0.027	37,900; 41,400
CGUG	1,830	0.027	68,600

The calculation of K for a nonbinding tetramer is simple since f is the same as for the trimer. Calculating K for a binding tetramer is more complicated for the reason discussed above in connection with Table VII. But here, the additional binding base can be either U or G so that a polymer composition cannot be selected that will ensure that in all cases the number of specific tetramer sites will be much larger than the number of sites to which the tetramer can only bind as a trimer.

To avoid the possibility that a tetramer would falsely appear to have a larger K than its parent trimer by dividing the same extent of binding $(R-1)$ by a much smaller f, the quantity $[(R-1)E_{res}]/A_{260}$ is tabulated in Table VIII. The tetramer f will be small when the additional binding base is G. In such cases the quantity $[(R-1)E_{res}]/A_{260}$ is indeed much larger for the tetramer than for the trimer, so that the calculated K is valid.

Since a tetramer can be considered as the addition of a binding base to the 5′ end of one trimer or to the 3′ end of another, the quantity $[(R-1)E_{res}]/A_{260}$ for the tetramer must be compared with the corresponding quantities for the binding of both trimers. In Table VIII the relevant data for the binding of the second trimer is given beneath the data for each binding tetramer. Since the tetramer can bind as both trimers, the quantity $[(R-1)E_{res}]/A_{260}$ should be much greater than the sum of the corresponding quantities for both trimers. It can be seen that this requirement is easi-

ly met except when one of the constituent trimers is GGU. In such cases $[(R - 1)E_{res}]/A_{260}$ for the tetramer is only two to three times larger than for the sum of the trimers. However, if the comparison is made to the binding of GGU, f for the tetramer is only slightly less than f for the trimer. Thus, the calculation of K is still valid.

In four out of eight cases, the addition of a nonbinding base did not change the value of K. However, the value of K was greatly decreased in one case and increased in the three other cases. An examination of these results shows that the pattern seen for adding a nonbinding base to GG is largely followed. All three cases in which the nonbinding base caused a substantial increase in K involved the addition of a pyrimidine. An additional characteristic of the effect of the nonbinding base is indicated by these results. No change in K was seen for any pair in which the nonbinding pyrimidine was added adjacent to a U in the binding sequence, but those cases involving adding a nonbinding pyrimidine adjacent to a G did exhibit an increase in K ranging from a factor of 1.6 to a factor of four.

By contrast, the addition of a binding base causes a very large increase in K. The increase in K varies from as little as a factor of three to as much as a factor of 450. Although the range of increase due to a nonbinding base overlaps the range of increase due to a binding base, the effect of adding a G or U to a particular trimer sequence is much larger than adding a nonbinding base to that sequence. Adding a U adjacent to a U in the trimer sequence causes a three- to eightfold increase in K compared to a negligible increase when a nonbinding base is added. If G is added adjacent to a G in the trimer sequence, the increase in K is large enough to dwarf the as much as fourfold increase in K caused by adding a nonbinding base. The only ambiguity arises when U is added to the 5′ end of a trimer adjacent to a G, in which case the increase in the value of K is no larger than the maximum increase caused by adding a nonbinding base.

Discussion

We have shown that only the Watson–Crick and G·G interactions are strong enough to cause oligonucleotide binding to synthetic polynucleotides. Thus, except for the oligonucleotides GG and GGG, any oligomer binding to an RNA must be binding to its Watson–Crick complementary sequence. A trimer (or tetramer) containing the sequence GG might be binding to a GG sequence if only the GG part of the entire trimer (or tetramer) sequence is binding to the RNA. However, if the entire oligomer sequence is bound, then it must be bound to its Watson–Crick complement.

To differentiate between the binding of only part of an oligonucleotide and the binding of the entire sequence, two approaches may be considered. First, this question could be answered if an expected value of K could be assigned to the binding of a given oligomer based on its length and G + C content. If it exhibited such a value of K binding to an RNA, it would be considered to be pairing its whole sequence. If it showed a much smaller K, it would be considered to be pairing only part of its sequence. Unfortunately, K can at least in some cases vary over a wide range with changes in the environment of the site, irrespective of any secondary structure. Therefore it is not possible to assign with sufficient precision a value of K that is to be expected for the binding of a given oligomer under all conditions in which it is pairing its entire sequence.

It does seem possible, however, to set some minimum values for the K to be expected for the binding of certain oligomers to an RNA, regardless of variations due to the conformation of the site. Of those tetramers capable of forming one G·C bond, the lowest K is observed for UUUG, which exhibits a K of 3000 when binding to poly($A_{4.6}$,C), a polymer that should contain sites in a favorable environment. The binding of UUUG to poly($C_{6.1}$,A) was not measured, but can be estimated from the data in Table IV to be about 2300. Of the tetramers capable of forming two G·C bonds, the one with the lowest K binding to poly($A_{4.6}$,C) is UUGG, with a K of 112,000. Its K binding to poly($C_{6.1}$,A) can be estimated to be about 25,000. Similarly, UGG is estimated to bind to poly($C_{6.1}$,A) with a K of about 2000. These considerations lead us to suggest lower limits of K = 1000, 3000, and 1000 for the tetramers of 25% G + C content, 50% G + C content, and for trimers of 67% G + C content, respectively. Any tetramer with a K less than these limits is presumed not to be binding its entire sequence. Any tetramer with a K greater than these limits will be classified by the criteria below.

The more promising approach is to compare the binding of the oligomer with the binding of a shorter oligomer whose sequence is contained in the sequence of the longer oligomer. If the longer oligomer binds much better than the shorter one, it is binding its entire sequence. If it binds only as well as the shorter oligomer, it is binding only part of its sequence. To test the reliability of this method, one must measure the effects of adding a binding vs. adding a nonbinding base to a short binding sequence. We have shown that a nonbinding base does not always leave unchanged the K of the shorter sequence. It may cause either a decrease or an increase, but only the latter case is of any concern. Since such an increase could be mistaken as indicating the specific binding of the entire oligomer sequence, it is necessary to demonstrate that the increase due to a nonbinding base can be differentiated from the increase due to a binding base. The only cases seen where a nonbinding base substantially increased K involved adding a nonbinding pyrimidine adjacent to a G in the binding oligomer. In such cases the increase in K was sometimes as much as a factor of four.

The increase in K seen upon the addition of a binding base was almost always larger than even the largest increase associated with the addition of a nonbinding base. One can arrive at quantitative estimates of the increase in K to be expected for adding various binding bases to different sequences. The association of two complementary sequences can be interpreted according to the model of Applequist and Damle (1965), in which the equilibrium constant for the formation of the first base pair is βs, and the equilibrium constant for the formation of each subsequent base pair is s. Since the inclusion of partially bonded states has quantitatively little effect upon the calculated K (Applequist and Damle, 1965; Martin, 1969), the all-or-none model can be used to give $K = \beta s^n$, where n is the number of base pairs formed. This equation is expected to apply to self-pairing oligomers. To consider the pairing of nonidentical sequences, the difference in stoichiometry can be absorbed into the initiation parameter β. To include the different stabilities of different base pairs and double-strand stacking interactions, s^n can be replaced by

$$\prod_{i=1}^{n} s_i$$

For our ad hoc treatment, we consider the binding constant of a trimer or tetramer to be a product of an initiation

constant ($\beta s_1 s_2$), which is equal to the K of the initiation dimer, and a product of constants ($s_3, s_3 s_4$) for pairing a particular base next to a particular base in an already formed helix, i.e., s is varied to account for different double-strand stacking interactions. The choice of the dimer sequence to be considered the initiation dimer is critical. It is assumed that the dimer sequence which has the largest K binding as an isolated dimer is the sequence with which the larger oligomer will initiate pairing. When calculating an s from the K's of two oligomers, it is essential that they be initiating identically. For example, one can compare the K's of GU and GUU to obtain s_{UU} because the K of GU is several orders of magnitudes higher than that of UU. Similarly one may compare the K's of GG and GGU since the K of GG is about 30 times that of GU. One may not compare the K's of GU and GUG since GU binds only three times better than UG so that GUG may be initiating quite differently from GU.

The values obtained for s_{UU}, s_{GU}, and s_{GG} calculated many different ways are listed in Table IX. The agreement is quite good. The aberrant values of $s_{UU} = 10$ and $s_{UG} = 4.0$ were both obtained using K for UGUU, so that the measured K for that oligomer may be too high by 30%. One also suspects the K of GGUU. It should be noted that these s values were all calculated using data from the binding to poly($A_{4.6}$,C). Before they can be applied with confidence to the binding of oligomers to RNAs, it will be necessary to study more polymers.

Having determined the relative effects of adding binding and nonbinding bases, we suggest the following guide lines for determining whether or not the entire sequence of an oligomer is binding.

(i) If the potential nonbinding base is an A, or if it is a pyrimidine adjacent to a U in the binding sequence, a longer sequence with a K less than 1.5 times the sum of the K's of the two constituent shorter sequences may be classed as nonbinding. A longer sequence with a K greater than five times this sum may be classed binding (2.5 times if the UG double-strand stack is involved). Any K between these limits probably represents binding of the entire sequence but only to a fraction of the RNA molecules, or to a site in which the residue in the site complementary to the added oligomer residue is in a conformation less favorable than the rest of the site. This latter situation may be expected to occur frequently in an RNA molecule where the single-strand regions are expected to be rather short so that the conformation of the single-strand region need not be uniform throughout its length. Such oligomers will be classed as intermediate.

(ii) For all other cases the respective limits should probably be placed at two and five. Here the intermediate category will also include those cases where the nonbinding base has a larger than usual effect (i.e., a two- to fourfold increase in K). Thus such occasional large effects of a nonbinding base may cause a nonbinding oligomer to be mislabeled as intermediate instead of nonbinding, but do not permit it to be labeled binding.

Table IX: Calculation of s_{UU}, s_{UG}, s_{GU}, and s_{GG}.[a]

Interaction	Oligomers Compared	s
UU	UUUU/UUUU	7.2
	UUG/UG	6.2
	UUUG/UUG	8.1
	UUGG/UGG	7.4
	GUU/GU	6.7
	GUUU/GUU	7.5
	GGUU/GGU	2.9
	UGUU/UGU	10
UG	UGG/GG	2.6
	UGGU/GGU	2.7
	UGU/GU	2.6
	UGUU/GUU	4.0
GU	GGU/GG	13
	UGGU/UGG	14
GG	GGG/GG	82

[a] All calculations use K's for the given oligomer binding to poly($A_{4.6}$,C).

Clearly much remains to be elucidated about the binding of oligomers to polymers. However, the experiments discussed here do point out the types of effects involved, and give an estimate of their magnitude.

References

Applequist, J., and Damle, V. (1965), *J. Am. Chem. Soc. 87,* 1450.

Comb, D. G., and Katz, S. (1963), *J. Biol. Chem. 238,* 3065.

Danchin, A., and Grunberg-Manago, M. (1970), *FEBS Lett. 9,* 327.

Grunberg-Manago, M. (1963), *Prog. Nucleic Acid Res. 1,* 93.

Hogenauer, G. (1970), *Eur. J. Biochem. 12,* 527.

Lewis, J. B. (1971), Ph.D. Thesis, Harvard University.

Lewis, J. B., and Doty, P. (1970), *Nature (London) 225,* 510.

Martin, F. H. (1969), Ph.D. Thesis, Harvard University.

Martin, F., and Uhlenbeck, O. C. (1971), *J. Mol. Biol. 57,* 201.

Michelson, A. M., Massouli, J., and Guschlbauer, W. (1967), *Prog. Nucleic Acid Res. 6,* 83.

Michelson, A. M., and Monny, C. (1967), *Biochim. Biophys. Acta 149,* 107.

Sander, C., and Ts'o, P. O. P. (1971), *J. Mol. Biol. 55,* 1.

Schimmel, P. R., Uhlenbeck, O. C., Lewis, J. B., Dickson, L. A., Eldred, E. W., and Schreier, A. A. (1972), *Biochemistry 11,* 642.

Tanford, C. (1961), Physical Chemistry of Macromolecules, New York, N.Y., Wiley, Chapter 4.

Uhlenbeck, O. C. (1969), Ph.D. Thesis, Harvard Universi-

Uhlenbeck, O. C. (1972), *J. Mol. Biol. 65,* 25.

Uhlenbeck, O. C., Baller, J., and Doty, P. (1970), *Nature (London) 225,* 508.